

Deep learning-based artificial intelligence for predicting risk and prognosis in patients with cardiovascular disease

Ki-Hyun Jeon^{1*}, Joon-myung Kwon^{2,3}, Kyung-Hee Kim¹, Jinsik Park^{1,3}

¹Department of Cardiology, Cardiovascular Center, Mediplex Sejong Hospital, Incheon, Republic of Korea

²Department of Emergency Medicine, Mediplex Sejong Hospital, Incheon, Republic of Korea

³Medical AI, Seoul, Republic of Korea

*Author for correspondence:

Ki-Hyun Jeon, Department of Cardiology, Cardiovascular Center, Mediplex Sejong Hospital, Incheon, Republic of Korea, Email: imcardio@gmail.com

Received date: October 14, 2020

Accepted date: October 28, 2020

Published date: November 04, 2020

Abstract:

Cardiovascular disease (CVD) is a major healthcare problem worldwide. Risk stratification and prognosis prediction are critical in identifying high-risk patients and in decision making to devise treatment strategies for patients with CVD. For this purpose, various models have been developed and validated against large amounts of population registry data by using conventional statistical methods such as regression-based models. However, these conventional models have a problem of over-generalization and are not applicable to all individual patients. Deep learning is a branch of artificial intelligence in which artificial neural networks are used to analyze data patterns; it is similar to functioning of the human neural system. An advantage of deep learning is the automatic learning of features and relationships from given data. Recently, deep learning achieved high performance in several medical domains, such as image classification, diagnosis, clinical outcome prediction, and gene analysis. The focus of this review is to summarize deep learning-based prediction models in patients with CVD in terms of accuracy in comparison with conventional models.

Keywords: Cardiovascular disease ■ Deep learning ■ Cardiac arrest ■ Acute myocardial infarction

Abbreviation: AF: Atrial Fibrillation; AI: Artificial Intelligence; ASCVD: Atherosclerotic Cardiovascular Disease; AUPRC: Area Under the Precision-Recall Curve; AUROC: Area Under the Receiver Operating Characteristic Curve; CVD: Cardiovascular Disease; CPR: Cardiopulmonary Resuscitation; ECG: Electrocardiography; ED: Emergency Department; EMS: Emergency Medical Service; NRI: Net Reclassification Index; NSTEMI: Non-ST Segment Elevation Myocardial Infarction; OHCA: Out-of-Hospital Cardiac Arrest; ROSC: Recovery of Spontaneous Circulation; STEMI: ST Segment Elevation Myocardial Infarction

Introduction

Cardiovascular Disease (CVD) is common in the general population and is a major healthcare problem worldwide [1]. Risk stratification and prognosis prediction to identify high-risk patients are essential in decision making to devise treatment strategies for patients with CVD. For this purpose, various risk-prediction models have been developed and validated against large amounts of population registry data by using conventional statistical methods such as regression-based models [2-4]. However, these conventional models use a limited number of pre-specified factors and have a problem of over-generalization, so call 'one size does not fit all'.

Deep learning is a subset of machine learning in which artificial neural networks are used to analyze different factors; it has a structure similar to that of the human neural system [5]. Recently, there have been considerable advancements in deep learning-based Artificial Intelligence (AI), enabling the use of qualified big data and enhanced computing power. In the medical field, deep learning algorithms have achieved state-of-the-art performance in, for example, the discrimination of medical images and diagnosis or prediction models, by overcoming the limitations of conventional statistical methods [6-9].

In this review, we summarize recent achievements in deep learning–based AI for risk stratification and prognosis prediction in patients with CVDs.

Risk Estimation and Prognosis Prediction Models in Cardiovascular Diseases

Various risk-prediction models have been developed for estimating the risk of an initial CVD event in individuals without a documented CVD and for predicting the prognosis of patients diagnosed with CVDs. These conventional prediction models are based on regression models such as the logistic model [10], Cox hazard model [2], and Weibull model [3]. Models were developed based on a large pool of representative datasets and several variables (usually established risk factors) to predict the probability of cardiovascular events, such as cardiac death, myocardial infarction, and stroke, in terms of the odds ratio, relative risks, or hazard ratio [11].

The most widely known conventional risk-prediction model is probably the Framingham risk score [2]. It was proposed in 1998 and was derived from a largely Caucasian population in Europe, including 2,489 men and 2,856 women, in the age range of 30–74 years at the time of the Framingham Heart Study examination from 1971 to 1974. In this model, the probability of cardiovascular events is calculated using statistical tests including age-adjusted linear regression, logistic regression to test for trends, and age-adjusted Cox proportional hazards regression as well as its accompanying C-statistic. The model for risk estimation of stroke in Atrial Fibrillation (AF), the CHA2DS2-VASc score, is widely used for decision making regarding the use of anticoagulants in patients with AF [12]. Other well-established and validated risk-prediction models include the Systematic Coronary Risk Evaluation (SCORE) CVD death risk score for the 10-year risk of a first fatal atherosclerotic event [3], American College of Cardiology/American Heart Association (ACC/AHA) Atherosclerotic Cardiovascular Disease (ASCVD) risk estimator for the 10-year risk of heart disease and stroke [13], and Thrombolysis In Myocardial Infarction (TIMI) risk score [14] or Global Registry of Acute Coronary Events (GRACE) score for mortality prediction in ST Segment Elevation Myocardial Infarction (STEMI) patients [15].

Although these conventional models based on regression were useful in clinical practice, these statistical methods use a limited number of predictive factors that operate in the same manner for all patients. In particular, these models assume constant effects of risk factors for different ages and levels of other risk factors. Therefore, these models have a problem of over-generalization and

are not applicable to all individual patients.

Artificial Intelligence, Machine Learning, and Deep Learning

AI is a term used to describe the application of computer science to simulate intelligent behavior and critical thinking for decision making in a manner comparable to a human being. The role of AI as a major catalyst in the healthcare revolution is unquestionable. AI has already started revolutionizing healthcare by leveraging big-data analysis to optimize healthcare services. It has made great advances in the field of disease diagnosis, drug development, personalized treatment, and improved gene editing [16-19]. Machine learning is an application of AI that provides computers the ability to automatically learn and improve from experience without explicit programming, human intervention, or assistance [20]. Deep learning is a type of machine learning that is inspired by the manner in which the human brain analyzes data. It combines computer science, statistics, and mathematical algorithms to find patterns and make decisions based on complex and big data. It includes feature learning, which is a set of methods that allows a model to be fed with raw data and to automatically identify the features and relationships needed for conducting a task [21,22]. It has been the most popular method for developing AI in recent years and has been empowered by big data and enhancements in computing power since 2010. In deep learning models, data are filtered through a cascade of multiple layers, with each successive layer using the output from the previous one to obtain its results [23,24].

Deep Learning–Based AI Algorithm for CVD Prediction

Recently, deep learning–based AI has achieved high performance in several medical domains, such as the detection of retinal disease, diagnosis of medical images, and electrocardiographic diagnosis of heart disease [25-28]. Predicting the risk and prognosis of CVD is a complex task with many factors to consider and needs time-consuming human operations to analyze that. Deep learning–based AI has excellent ability to solve problems automatically by analyzing these complex factors.

CVD risk-prediction model

Prediction models to estimate the risk of developing CVD require large amounts of complex data that are suitable for the application of deep learning. Cho et al. [29] developed a CVD prediction model using large-scale cohort data including 412,030 Korean adults in the National Health Insurance Service-Health Screening Cohort (NHIS-HEALS) for internal validation, 178,875 adults in the National Health Insurance Service–National Sample Cohort (NHIS-NSC) for the first external validation, and the 4,296 European adults in the Rotterdam Study [30] for the second external validation. In the external validation based on the Rotterdam Study, which included participants of different

ethnicities, the model demonstrated a C-statistic of 0.860 (0.824–0.897) in men and 0.867 (0.830–0.903) in women, as well as improved reclassification compared with conventional Cox regression (net reclassification index [NRI] of 36.9% in men and 31.8% in women).

Prediction model for mortality and prognosis of cardiac arrest

Cardiac arrest is a catastrophic event that leads to sudden cardiac death. It affects not only patients with CVD but also healthy people. Among patients with Return Of Spontaneous Circulation (ROSC), the in-hospital mortality rate is 50%-70%, and the majority sustain ischemic neurological injury [31,32].

Prediction of in-hospital cardiac arrest

Various medical conditions can cause in-hospital cardiac arrest, and the survival discharge rate of these patients is less than 20% [33]. More than half of in-hospital cardiac arrest cases result from respiratory failure or hypovolemic shock, and 80% of the patients who experienced cardiac arrest showed signs of deterioration in the eight hours before the cardiac arrest [34,35]. For predicting cardiac arrest, several risk-score models based on vital signs, including blood pressure and heart rate, are used for patient safety [36]. Kwon et al. [37] developed a deep learning-based early warning system (DEWS) to predict in-hospital cardiac arrest using vital sign data, including systolic blood pressure, heart rate, respiratory rate, and body temperature, from 52,131 patients. The area under the receiver operating characteristic curve (AUROC) of 0.850 and area under the precision-recall curve (AUPRC) of 0.044 for DEWS were significantly higher than those of a modified early warning score (AUROC of 0.603 and AUPRC of 0.003), which is one of the most widely used conventional approaches. An interesting point in this study is that the risk score of cardiac arrest in DEWS increased from 24 h before the cardiac arrest, and a DEWS >50% was found in patients 14 h before the cardiac arrest.

Prediction of mortality and neurologic outcomes for out-of-hospital cardiac arrest

Out-of-hospital cardiac arrest (OHCA) is a leading cause of global mortality. Globally, the percentage of survival to discharge was 5%–10%, depending on the region [38]. Even with successful resuscitation, ischemic neurological damage is inevitable in many cases. The prediction of prognosis is important in decision making to devise treatment strategies for patients with OHCA. Kwon and Jeon et al. [39] developed and validated a deep learning-based out-of-hospital cardiac-arrest prognostic system (DCAPS) for predicting neurologic recovery and survival to discharge. The deep

learning model was developed using data from the Korea OHCA Registry (KOHCAR) in South Korea, in which 36,190 patients with OHCA from 712 emergency departments (EDs) were enrolled (Figure 1) [39]. As predictor variables during model development, the authors utilized only the information available at the time of ROSC, including age, sex, place of OHCA, etiology of OHCA (disease or trauma), ROSC in emergency medical service (EMS), event witness, bystander cardiopulmonary resuscitation (CPR), initial electrocardiography (ECG) rhythm of EMS, initial ECG rhythm of ED, and time from ED visit to ROSC. The AUROC of DCAPS was 0.953 (95% confidence interval [CI], 0.952–0.954) for neurologic recovery, which is significantly higher than the value of 0.817 achieved using the conventional model (95% CI, 0.815–0.820). In terms of survival to discharge, the AUROC of the DCAPS and conventional model were 0.901 (95% CI, 0.900–0.903) and 0.736 (95% CI, 0.734–0.739) (Figure 2) [43]. Therefore, deep learning-based AI model accurately predicted the neurologic recovery and survival to discharge of OHCA patients, outperforming the conventional method.

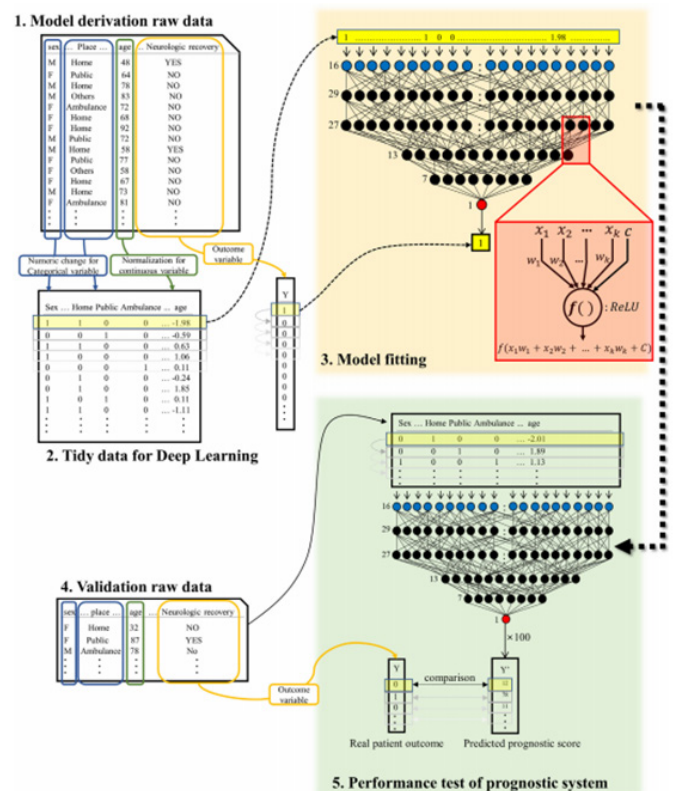
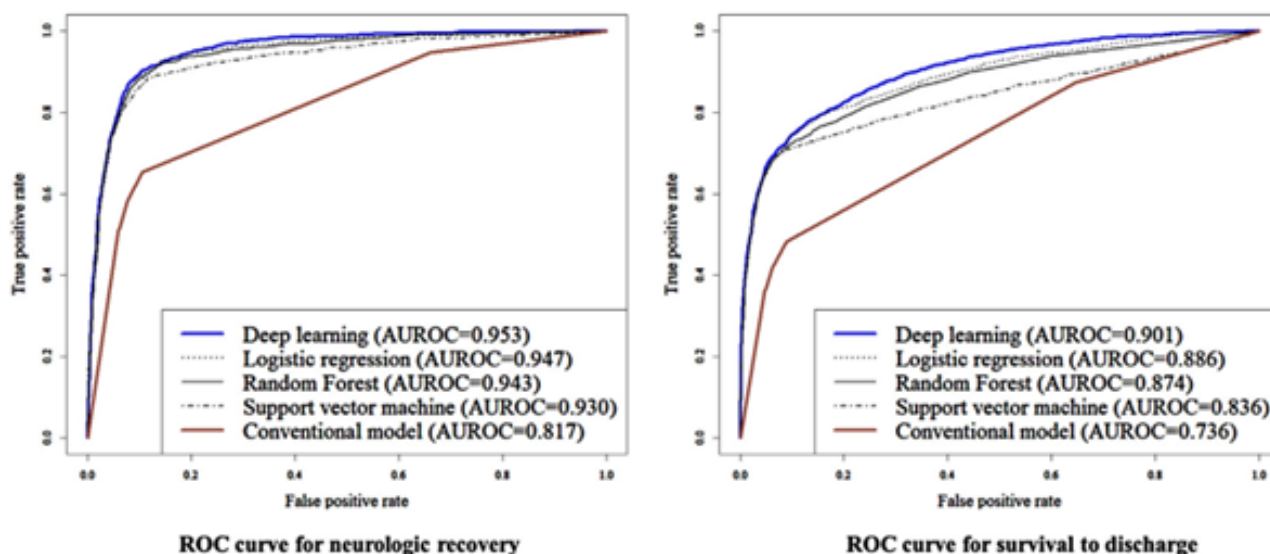


Figure 1: Development and validation of deep learning-based prognostic model. Reprinted from Kwon et al., Resuscitation 139: 84-91 (2019), Copyright by Elsevier



Prognostic model	Neurologic recovery			Prognostic model	Survival to discharge		
	AUROC	(95% CI)	<i>p</i> -value†		AUROC	(95% CI)	<i>p</i> -value†
Deep learning (DCAPS)	0.953	(0.952–0.954)	–	Deep learning (DCAPS)	0.901	(0.900–0.903)	–
Logistic regression	0.947	(0.946–0.948)	<0.001	Logistic regression	0.886	(0.884–0.888)	<0.001
Random forest	0.943	(0.942–0.945)	<0.001	Random forest	0.874	(0.872–0.876)	<0.001
Support vector machine	0.930	(0.929–0.932)	<0.001	Support vector machine	0.836	(0.833–0.838)	<0.001
Conventional model	0.817	(0.815–0.820)	<0.001	Conventional model	0.736	(0.734–0.739)	<0.001

Figure 2: ROC curve and AUROC for neurological recovery and survival to discharge. AUROC, area under receiver operating characteristic curve; CI, confidence interval; DCAPS, deep learning–based out-of-hospital cardiac arrest prognostic score. †The alternative hypothesis for this *p*-value is that there is a difference the between the area under the curve of deep learning (DCAPS) and those of other methods. Reprinted from Kwon et al., Resuscitation 139: 84-91 (2019). Copyright by Elsevier

Prediction of mortality in patients with AMI

Many efforts have been made to accurately predict the prognosis of patients with acute myocardial infarction (AMI). Conventional risk scoring systems, including TIMI [10,14], GRACE [15], and the acute coronary treatment and intervention outcomes network (ACTION) [40], are widely validated and accepted scores that are estimated using patients’ clinical information. However, these prognostic models have limitations in current daily practice. Firstly, these systems are questionable in contemporary practice because they had been developed 20 years ago. Additionally, as these models use only selective variables based on a conventional statistical method, there is a possibility of loss of important information. Shouval et al. [41] developed a machine-learning algorithm to predict 30-day mortality following STEMI and compared it with the conventional GRACE and TIMI scoring systems. The best accuracy achieved using the machine-learning algorithm (AUC of 0.91 and standard deviation [SD] of 0.04) was similar to that of GRACE (AUC of 0.87 and SD of 0.06) and better than that of TIMI (AUC of 0.82, SD of 0.06, and *P*<0.05). In a study on deep learning–based risk stratification for mortality of patients with

AMI (DAMI) [42], a deep learning algorithm was developed from 22,875 AMI patients from the Korean working group of acute myocardial infarction (KorMI) registry. The algorithm used a total of 37 variables of demographic and laboratory data. During the accuracy test for STEMI patients, the AUC of DAMI was 0.905 (95% CI, 0.902–0.909), and this result significantly outperformed the GRACE score (0.851 with 95% CI, 0.846–0.856), ACTION score (0.852 with 95% CI 0.847–0.857), and TIMI score (0.781 with 95% CI 0.775–0.787). The algorithm also showed better accuracy than conventional models in a non-ST segment elevation myocardial infarction (NSTEMI) patient group. Figure 3 shows the results of reclassification of individuals predicted to be in the intermediate-risk group through additional DAMI assessment [39]. Among the 3,526 patients who were placed in the intermediate-risk group based on the GRACE score, 1937 patients were reclassified into the low-risk group. Furthermore, among 50 patients who met with in-hospital death, 24 patients were reclassified into the high-risk group based on the DAMI score. This implies that DAMI can differentiate the mortality risk of patients with AMI more sensitively than conventional models.

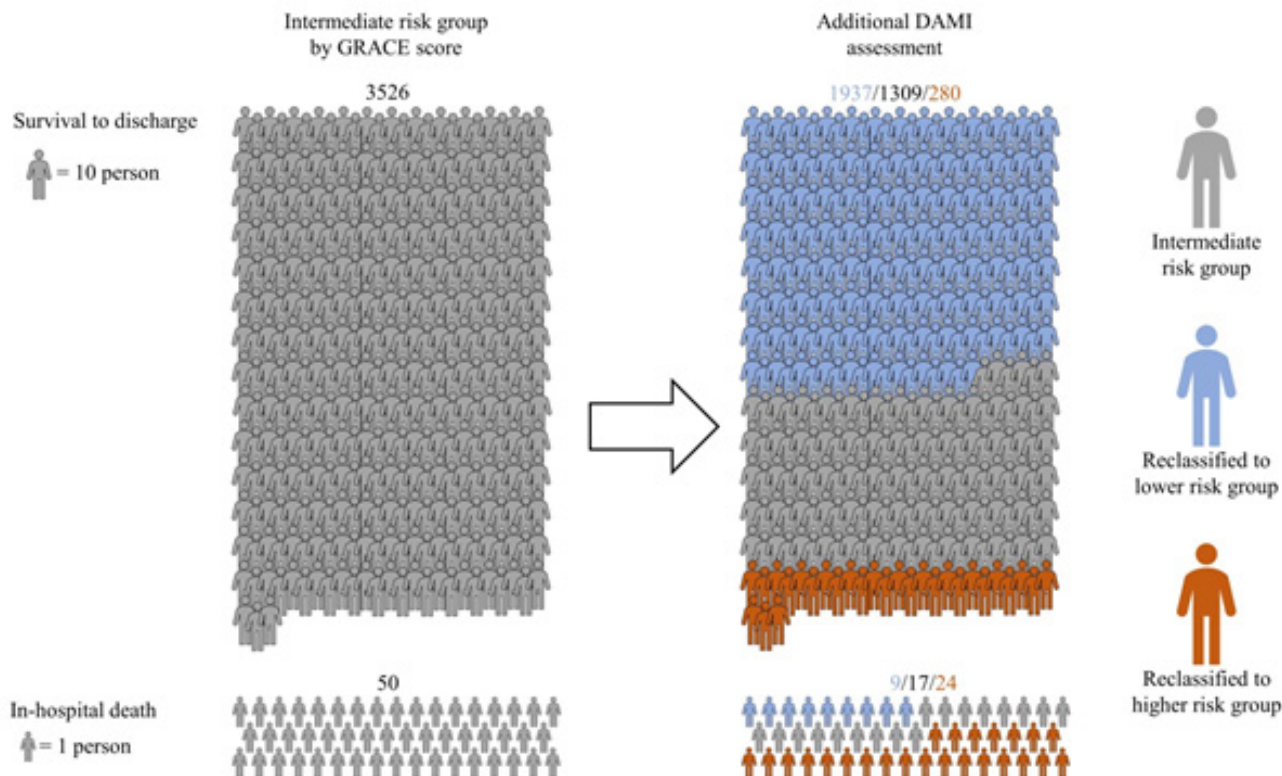


Figure 3: Reclassification of individuals predicted to be in the intermediate-risk group through additional DAMI assessment. GRACE, Global Registry of Acute Coronary Events; DAMI, deep learning–based risk stratification for mortality of patients with AMI. Reprinted from Kwon et al., PLoS One. 14: e0224502 (2019).

Limitation of Deep Learning

AI based on deep learning has enormous potential in the medical field, and it could improve the accuracy of diagnosis and support clinical decisions for many diseases. However, it is also necessary to clearly recognize the limitations of AI and make efforts to overcome them. One of the most important characteristics of deep learning is that it does not use any medical knowledge; rather, it uses only the relationship among variables of given data and can easily overfit the dataset. Therefore, external validation is essential in medical AI research. Here, the term “external” refers not only to exclusively separated dataset, but also to data from different environments. Validation of an established AI model in a completely new dataset is mandatory for AI research to overcome overfitting. For that, cross-validation is one of the preferred methods to reduce the variance in prediction error and to give an insight on how the model will generalize to an independent dataset [22]. The second limitation is that current AI technology cannot reveal the decision process of deep learning, which is so-called a black box. In other words, although we can develop deep learning–based AI by fitting each coefficient, we cannot interpret the AI in terms of its approach to decision making. Interpretable deep learning has been studied recently, but it takes much time to apply the theory to the medical field. As with past medical research, studies on AI must not only provide accurate results but also attempt to analyze

and understand AI.

Conclusion

We are currently on the brink of the “fourth industrial revolution,” a technological revolution that will fundamentally change the patterns by which we live, work, and relate to one another. Deep learning has achieved state-of-the-art performance in several medical domains and outperforms existing conventional methods. Deep learning showed better performance than conventional models in risk stratification and prognosis prediction for CVD and could be of great help in the evaluation and treatment of patients with CVD in the future.

Conflict of Interest

The authors declare no conflict of interest.

Funding

This study received no funding.

Author’s contribution

Ki-Hyun Jeon and Joon-myung Kwon contributed equally to this article.

References

- Roth GA, Johnson C, Abajobir A et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol.* 70(1): 1-25 (2017).
- D'Agostino RB, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: Results of a multiple ethnic group's investigation. *JAMA.* 286(2): 180-7 (2001).
- Conroy RM, Pyorala K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *Eur Heart J.* 24(11): 987-1003 (2003).
- Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *BMJ.* 335: 136 (2007).
- Krittanawong C, Johnson KW, Rosenson RS, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J.* 40(25): 2058-2073 (2019).
- Bizopoulos P, Koutsouris D. Deep learning in cardiology. *IEEE Rev Biomed Eng.* 12: 168-93 (2019).
- Kwon JM, Kim KH, Jeon KH, et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PLoS One.* 14(7): e0219302 (2019).
- Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J Am Coll Cardiol.* 73: 1317-35 (2019).
- Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology.* 290(1): 218-28 (2019).
- Morrow DA, Antman EM, Charlesworth A, et al. TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation: an intravenous nPA for treatment of infarcting myocardium early II trial substudy. *Circulation.* 102(17): 2031-7 (2000).
- Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J.* 38: 1805-14 (2017).
- January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *Circulation.* 130: 2071-104 (2014).
- Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 63: 2935-59 (2014).
- Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. *JAMA.* 284(7): 835-42 (2000).
- Eagle KA, Lim MJ, Dabbous OH, et al. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month post-discharge death in an international registry. *JAMA.* 291(22): 2727-33 (2004).
- Teng SY, Yew GY, Sukacova K, et al. Microalgae with artificial intelligence: A digitalized perspective on genetics, systems and products. *Biotechnol Adv.* 44: 107631 (2020).
- Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med.* 25: 44-56 (2019).
- Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: Review, opportunities and challenges. *Brief Bioinform.* 19(6): 1236-1246 (2018).
- Krittanawong C, Virk HUH, Bangalore S, et al. Machine learning prediction in cardiovascular diseases: A meta-analysis. *Sci Rep.* 10: 16057 (2020).
- Deo RC. Machine learning in medicine. *Circulation.* 132: 1920-30 (2015).
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 521: 436-44 (2015).
- Kagiyama N, Shrestha S, Farjo PD, et al. Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. *J Am Heart Assoc.* 8(17): e012788 (2019).
- Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. *PNAS.* 20: 1907373 (2020).
- Benjamins JW, Hendriks T, Knuuti J, et al. A primer in artificial intelligence in cardiovascular medicine. *Neth Heart J.* 27: 392-402 (2019).
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 316: 2402-10 (2016).
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 542: 115-8 (2017).
- Kwon JM, Jeon KH, Kim HM et al. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace.* 22(3): 412-9 (2020).
- Kwon JM, Kim KH, Jeon KH et al. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J.* 49: 629-39 (2019).
- Cho IJ, Sung JM, Kim HC et al. Development and external validation of a deep learning algorithm for prognostication of cardiovascular outcomes. *Korean Circ J.* 50: 72-84 (2020).
- Hofman A, Brusselle GG, Murad SD, et al. The Rotterdam Study: 2016 Objectives and design update. *Eur J Epidemiol.* 30: 661-708 (2015).
- Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart disease and stroke statistics-2017 update: A report from the American Heart Association. *Circulation.* 135(10): e146-e603 (2017).
- Grasner JT, Lefering R, Koster RW, et al. Corrigendum to "EuReCa ONE-27 Nations, ONE Europe, ONE Registry: a prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe". *Resuscitation.* 109: 145-6 (2016).
- Nadkarni VM, Larkin GL, Peberdy MA, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. *JAMA.* 295(1): 50-7 (2006).
- Schein RM, Hazday N, Pena M, et al. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest.* 98(6): 1388-92 (1990).
- Franklin C, Mathew J. Developing strategies to prevent in hospital cardiac arrest: Analyzing responses of physicians and nurses in the hours before the event. *Crit Care Med.* 22(2): 244-7 (1994).

Review Article

36. Jones DA, DeVita MA, Bellomo R. Rapid-response teams. *N Engl J Med.* 365: 139-46 (2011).
37. Kwon JM, Lee Y, Lee Y, et al. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *J Am Heart Assoc.* 7: e008678 (2018).
38. Myat A, Song KJ, Rea T. Out-of-hospital cardiac arrest: Current concepts. *Lancet.* 391: 970-9 (2018).
39. Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. *Resuscitation.* 139: 84-91 (2019).
40. McNamara RL, Kennedy KF, Cohen DJ, et al. Predicting in-hospital mortality in patients with acute myocardial infarction. *J Am Coll Cardiol.* 68: 626-35 (2016).
41. Shouval R, Hadanny A, Shlomo N, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infarction: An Acute Coronary Syndrome Israeli Survey data mining study. *Int J Cardiol.* 246: 7-13 (2017).
42. Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. *PLoS One.* 14(10): e0224502 (2019).